

Ab Initio Geometry Optimization for Large Molecules

FRANK ECKERT, PETER PULAY,* HANS-JOACHIM WERNER

Institut für Theoretische Chemie, Universität Stuttgart, Pfaffenwaldring 55, D-70569 Stuttgart, Germany

Received 19 November 1996; accepted 15 February 1997

ABSTRACT: Various geometry optimization techniques are systematically investigated. The rational function (RF) and direct inversion in the iterative subspace (DIIS) methods are compared and optimized for the purpose of geometry optimization. Various step restriction and line search procedures are tested. The model Hessian recently proposed by Lindh et al. has been used in conjunction with different Hessian update procedures. Optimization for over 30 molecules have been performed in Z-matrix coordinates, local normal coordinates, and curvilinear natural internal coordinates, using the same approximations for the Hessian in all cases. The most effective and stable procedure for optimization of equilibrium structures was found to be the DIIS minimization in natural internal coordinates using the BFGS update of the model Hessian. Our method shows faster overall convergence than all previously published methods for the same test suite of molecules. © 1997 John Wiley & Sons, Inc. *J Comput Chem* **18**: 1473–1483, 1997

Keywords: geometry optimization; DIIS optimization algorithm; Hessian; natural internal coordinates; rational function

Introduction

One of the most important areas of applied quantum chemistry is the determination of stationary points (equilibrium and transition struc-

tures) on potential energy surfaces. The availability of *ab initio* gradient techniques^{1–3} has made the development of efficient geometry optimization methods possible (for reviews see Refs. 4–6). For larger systems, *ab initio* geometry optimization is only practical using analytic gradients and approximate Hessians employing some variant of the quasi-Newton algorithm.⁷

The efficiency of an optimization, measured by the number of energy and gradient evaluations needed to achieve convergence, depends mainly on three factors: (1) the optimization algorithm controlling the step size and direction; (2) the ap-

* Permanent address: Department of Chemistry and Biochemistry, University of Arkansas, Fayetteville, AR 72701, USA

Correspondence to: H.-J. Werner

Contract/grant sponsor: Air Force Office of Research

Contract/grant sponsor: National Science Foundation

Contract/grant sponsor: Alexander-von-Humboldt Foundation

proximation to the Hessian; and (3) the coordinates used to describe the system. Reasonable starting geometries can easily be obtained by graphics-based molecular model builders with subsequent preoptimization by molecular mechanics force-field methods.

The most important algorithms for *ab initio* geometry optimization are the rational function (RF) method,⁸ the closely related eigenvector following (EF) algorithm,⁹ and the direct inversion in the iterative subspace (DIIS) method.¹⁰ There are many possibilities for approximating the Hessian.⁴ Widely used are approximations based on empirical force fields, such as valence force fields,^{11,12} or molecular mechanics force fields.¹³ Recently, a very simple model Hessian has been proposed by Lindh et al.,¹⁴ which depends on the actual geometry and has been shown to be at least as effective as much more elaborate force fields. This model Hessian has been employed in the present work, together with various Hessian update procedures.

During recent years much of the attention has shifted toward the coordinate system in which the optimization is carried out. The simplest choice is Cartesian coordinates,^{13,15} but these are most strongly coupled and work well only if good approximations to the Hessian are available. To reduce the couplings various internal coordinates have been suggested: curvilinear natural internal coordinates^{16,17}; redundant internal coordinates^{18,19}; and the delocalized internal coordinates by Baker et al.²⁰

In the present article, we compare optimizations in local normal coordinates (defined by the eigenvectors of the approximate Hessian) with curvilinear internal coordinates.^{16,17} For comparison, some calculations are also performed in Z-matrix coordinates. For purposes of unambiguous comparison the model Hessian has been transformed into any of the three different coordinate systems studied. Furthermore, the same optimization methods (RF or DIIS) were used for each choice of coordinates.

The following section summarizes the optimization algorithms, coordinates, and Hessian approximations employed in this work. In a later section these methods are systematically tested for a wide variety of molecules. The RF and DIIS methods are found to perform about equally well when combined with the model Hessian of Lindh et al.¹⁴ Natural internal coordinates^{16,17} were found to be more effective than Z-matrix and local normal coordinates.

Methodology

RATIONAL FUNCTION METHOD

For a surface of n independent coordinates, a quadratic Taylor expansion can be used to approximate the energy surface in the neighborhood of a point $\mathbf{x}_k = (x_1, \dots, x_n)$:

$$E^{(2)}(\mathbf{x}) = E(\mathbf{x}_k) + \mathbf{g}_k^\dagger \mathbf{s}_k + \frac{1}{2} \mathbf{s}_k^\dagger \mathbf{H}_k \mathbf{s}_k \quad (1)$$

The step vector $\mathbf{s}_k = (\mathbf{x} - \mathbf{x}_k)$ describes a displacement from the reference geometry \mathbf{x}_k ; \mathbf{g}_k and \mathbf{H}_k are the gradient vector and Hessian matrix (first and second derivative of the energy) at \mathbf{x}_k ; the subscript k refers to the step number. Applying the stationarity condition, $\partial E^{(2)} / \partial \mathbf{s}_k = 0$, to eq. (1) leads to the quasi-Newton (QN) step:

$$\mathbf{s}_k = -\mathbf{H}_k^{-1} \mathbf{g}_k \quad (2)$$

The QN step is only reasonable if the Hessian has the correct eigenvalue structure (i.e., for minimizations all eigenvalues must be positive). If the Hessian is not positive definite, or the QN step is too large, one can modify eq. (1) by adding a term $\frac{1}{2} \lambda \mathbf{s}_k^\dagger \mathbf{s}_k$. Minimization of the resulting functional leads to a level shift in the denominator:

$$\mathbf{s}_k = -(\mathbf{H}_k + \lambda \mathbf{1})^{-1} \mathbf{g}_k \quad (3)$$

A sufficiently large value of λ holds $(\mathbf{H}_k + \lambda \mathbf{1})$ positive definite and restricts the optimization step to a trust region in which the quadratic approximation to the energy surface is reasonable.

A natural choice of λ is provided by the rational function (RF) method.^{8,9} In the simplest version of this method, the linear equation, eq. (3), is replaced by an eigenvalue equation:

$$\begin{pmatrix} \mathbf{H}_k & \mathbf{g}_k \\ \mathbf{g}_k^\dagger & 0 \end{pmatrix} \begin{pmatrix} \mathbf{s}_k \\ 1 \end{pmatrix} = \lambda_k \begin{pmatrix} \mathbf{s}_k \\ 1 \end{pmatrix} \quad (4)$$

The matrix on the left-hand side of eq. (4) is the *augmented Hessian* matrix of dimension $n + 1$. Resolving eq. (4) yields:

$$\lambda_k = \mathbf{g}_k^\dagger \mathbf{s}_k \quad (5)$$

$$\mathbf{g}_k + (\mathbf{H}_k - \lambda_k \mathbf{1}) \mathbf{s}_k = 0 \quad (6)$$

Thus, eq. (6) is equivalent to eq. (3) with λ given by eq. (5). Note that λ_k is negative in general, as the step direction is roughly opposite to the direction of the gradient.

For minimizations, λ_k is chosen as the lowest eigenvalue of the augmented Hessian matrix. The corresponding eigenvector gives, after scaling its last element to 1, the optimization step. This choice of λ_k ensures that $(\mathbf{H}_k - \lambda \mathbf{1})$ is positive definite, but it does not guarantee convergence. If the Hessian has small eigenvalues, the RF step may become very large and overshoot the minimum. It is then necessary to restrict the step size. There are several ways of doing this.

- (i) Scaling of the whole step vector if its norm exceeds a given threshold ($|\mathbf{s}_k| < smax$).
- (ii) Restricting the values of each component to a maximum value ($smax$).
- (iii) A more sophisticated *dynamical step scaling* procedure is obtained by including the constraint $|\mathbf{s}_k| < smax$ in the minimization. This leads to a modified augmented Hessian⁸:

$$\begin{pmatrix} \mathbf{H}_k/\alpha & \mathbf{g}_k \\ \mathbf{g}_k^\dagger & 0 \end{pmatrix} \begin{pmatrix} \mathbf{s}_k \\ 1/\alpha \end{pmatrix} = \lambda_k \begin{pmatrix} \mathbf{s}_k \\ 1/\alpha \end{pmatrix} \quad (7)$$

where α determines the Lagrange multiplier λ_k :

$$\lambda_k = \alpha \mathbf{g}_k^\dagger \mathbf{s}_k \quad (8)$$

$$\mathbf{g}_k + (\mathbf{H}_k - \alpha \lambda_k \mathbf{1}) \mathbf{s}_k = 0 \quad (9)$$

If the length of the step vector exceeds the threshold $smax$ then α is increased and the step is recalculated from eq. (7) until the condition $|\mathbf{s}_k| \leq smax$ is fulfilled. In contrast to a simple scaling, this method also modifies the direction of the step vector. For large α it approaches the steepest descent method. Thus, convergence is guaranteed by an appropriate choice of $smax$. This "step restricted augmented Hessian method"⁷ has been used successfully in MCSCF optimization techniques.^{21,22} We find, however (see "Limiting the Optimization Step" subsection), that, despite its theoretical justification, this method does not improve the convergence of geometry optimization in our examples. In fact, the simplest method (i.e., restricting all individual components of the step vector \mathbf{s}_k to a maximum value) usually leads to best overall convergence.

LINE SEARCH

An alternative way of improving the stability of optimization is to accept only the direction of the

step and determine the step length by locating the minimum between the current geometry \mathbf{x}_k and the previous geometry $\mathbf{x}_{k-1} = \mathbf{x}_k - \mathbf{s}_{k-1}$. For approximately quadratic surfaces, such a line search procedure is valid if the scalar products $\mathbf{g}_k^\dagger \mathbf{s}_{k-1}$ and $-\mathbf{g}_{k-1}^\dagger \mathbf{s}_{k-1}$ are both negative (i.e., a minimum exists on the line between the two points). *Exact line search* procedures minimize the energy along \mathbf{s}_{k-1} . Due to the numerous energy calculations needed, such procedures are not efficient in *ab initio* calculations. *Partial line search* procedures try to interpolate the energy minimum between the two points. The algorithm examined in this work is based on a proposal by Schlegel.²³ It seeks the minimum along a quartic polynomial built from the energies and gradients at \mathbf{x}_k and \mathbf{x}_{k-1} with the auxiliary condition $\partial^2 E / \partial x^2 \geq 0$ at both points, to insure that the polynomial has only one minimum. The geometry, energy, and gradient interpolated in this manner are used to determine the new optimization step \mathbf{s}_k .

GEOMETRY DIIS

The DIIS method¹⁰ attempts to find an optimum linear combination of the previous geometries. The basic assumption of the DIIS method is that the error vector, \mathbf{e}^* , which is a measure for the deviation of the actual geometry from the minimum, is a linear function of the geometry, and thus can be approximated by a linear combination of the error vectors, $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_k$, obtained in the previous iterations $1 \dots k$:

$$\mathbf{e}^* = \sum_{i=1}^k c_i \mathbf{e}_i \quad (10)$$

As discussed in what follows, several choices are possible for vectors \mathbf{e}_i (e.g., the gradient). The best linear combination of geometries is determined by minimizing the norm of the interpolated error vector \mathbf{e}^* with the auxiliary condition $\sum_i c_i = 1$. Using the Lagrangian method of constrained optimization yields an inhomogeneous system of $k+1$ linear equations for coefficients c_i :

$$\begin{pmatrix} \mathbf{A} & \mathbf{1} \\ \mathbf{1}^\dagger & 0 \end{pmatrix} \begin{pmatrix} \mathbf{c} \\ -\lambda \end{pmatrix} = \begin{pmatrix} \mathbf{0} \\ 1 \end{pmatrix} \quad (11)$$

where $A_{ij} = \mathbf{e}_i^\dagger \mathbf{e}_j$ is the DIIS matrix built from scalar products of the error vectors, $\mathbf{1}$ is a vector of length k with all elements 1, and λ is the La-

grangian multiplier. The linear equations may be conveniently solved by diagonalization of the DIIS matrix and subsequent renormalization of the coefficient vector \mathbf{c} , such that $\sum_i c_i = 1$. By monitoring its eigenvalues, one can detect (approximate) linear dependencies in the DIIS matrix \mathbf{A} (\mathbf{A} is scaled such that the last diagonal element becomes unity). In the case of linear dependency, the largest error vectors (i.e., the most remote from the actual geometry) are successively removed from the DIIS matrix until \mathbf{A} becomes reasonably well conditioned. The coefficients c_i obtained from eq. (11) are used to interpolate the geometry and gradient of the system:

$$\mathbf{x}_k^* = \sum_{i=1}^k c_i \mathbf{x}_i \quad (12)$$

$$\mathbf{g}_k^* = \sum_{i=1}^k c_i \mathbf{g}_i \quad (13)$$

The new geometry is then obtained by taking a quasi-Newton optimization step using the interpolated geometry and gradient:

$$\mathbf{x}_{k+1} = \mathbf{x}_k^* - \mathbf{H}_k^{-1} \mathbf{g}_k^* \quad (14)$$

Alternatively, geometry relaxation can be achieved by performing an RF step as described previously using \mathbf{x}_k^* and \mathbf{g}_k^* . The QN or RF relaxation steps $\mathbf{s}_k^* = \mathbf{x}_{k+1} - \mathbf{x}_k^*$ were scaled if their norm exceeded 0.3 (measuring the bond lengths in a_0 and angles in radians). Additionally, the individual components of the entire GDIIS step (i.e., interpolation plus relaxation), $\mathbf{s}_k = \mathbf{x}_{k+1} - \mathbf{x}_k$, were restricted to a maximum value (0.3 a_0 or radians).

The convergence properties of the GDIIS optimization depend on the choice of error vectors that determine the DIIS matrix \mathbf{A} . In the original work by Császár and Pulay,¹⁰ the error vectors were approximated by quasi-Newton steps built from the approximate Hessian matrix \mathbf{H}_k and the previous gradients \mathbf{g}_i of the optimization, $\mathbf{e}_i = -\mathbf{H}_k^{-1} \mathbf{g}_i$ and $A_{ij} = \mathbf{g}_i^\dagger \mathbf{H}_k^{-1} \mathbf{H}_k^{-1} \mathbf{g}_j$, corresponding to an interpolation in the subspace of the optimization steps (geometries). In practice, we found that optimizations using this interpolation scheme showed oscillative behavior in some of the test cases and thus converged slowly. Alternative approaches are DIIS interpolations that attempt to minimize the energy (i.e., $A_{ij} = \mathbf{g}_i^\dagger \mathbf{H}_k^{-1} \mathbf{g}_j$) or the gradients (i.e., $A_{ij} = \mathbf{g}_i^\dagger \mathbf{g}_j$). One would expect the gradient interpolation to perform better with stiff molecules, because strong bonds that have large gradient

components are weighed strongly, whereas the interpolation of the steps or the energy should perform better with flexible molecules. We also tested an approximate energy interpolation scheme using the diagonal elements of the Hessian restricted to $0.5 \geq H_{ii} \geq 3.0$ aJ/Å. The purpose of this restriction is to prevent the interpolation from weighing torsional motions with small force constants too much, and stiff bonds too little. For most cases we find that the simplest method, using the gradient for the error vector yields the fastest convergence (cf. next section).

HESSIAN MATRIX

Another highly important factor for the convergence behavior of geometry optimizations is the approximation to the Hessian matrix, which determines the harmonic (second order) coupling between the coordinates. A good approximation to the Hessian matrix will accelerate the convergence of any optimization. Calculation of the exact analytical or numerical Hessian is computationally costly [$\sim O(n)$ times the energy and gradient calculations] and is usually not economical in *ab initio* optimizations on ground-state molecules.

For the latter, the usual procedure is to calculate an approximate Hessian from a molecular force field and improve it by an updating procedure. A promising alternative to this approach is the "model Hessian" recently proposed by Lindh et al.¹⁴ The model Hessian is a very simple forcefield approximation derived from only 15 parameters, which incorporates quadratic contributions from bond stretchings, angle bending, and torsional angle bending. The model Hessian depends on the geometry and is recalculated in every optimization step. As shown in Ref. 14, it outperforms better approximations to the Hessian, which are calculated only once at the start of the optimization. In addition, it is useful to apply updating to the model Hessian.

Updating procedures try to improve an approximate Hessian matrix using the geometries and gradients generated during the optimization history (for a review see Ref. 4). Two methods are widely used for energy minimizations: the BFGS update²⁴⁻²⁷ and the conjugate gradient update,²⁸ as modified by Schlegel.²³ The latter algorithm excludes points from the update if they are too far from the actual point or if the geometries are linearly dependent, avoiding updates with "chemically insignificant" geometries.²³ It also reorders the optimization's history by decreasing energy.

The BFGS update uses just the last m geometries and gradients. The BFGS method has the tendency to keep the Hessian positive definite, which is very useful in minimizations, because negative Hessian eigenvalues may lead to uphill steps on the PES.

COORDINATE SYSTEMS

The choice of the coordinate system is an important factor for the convergence of geometry optimizations. For strictly quadratic surfaces (i.e., very close to the minimum), all nonsingular coordinate systems are equivalent, provided that the Hessian is appropriately transformed. Any optimization algorithm will run into difficulties in the presence of large anharmonic (cubic and higher) couplings. Simple Hessian estimates (e.g., a diagonal one) work only for weakly coupled coordinates. The coordinate system should therefore be chosen such that coupling terms in the potential function between coordinates are minimized.

Cartesian coordinates, widely used in molecular mechanics and modeling programs, are the simplest choice, but they are highly coupled and also include redundant translations and rotations. Curvilinear constraints can be imposed only approximately in Cartesian optimizations.^{29,30} However, if a good approximation to the Hessian is available, Cartesians work well, as shown by Baker^{13,15} and as expected from the previous discussion.

Internal coordinates are more appealing from the chemist's point of view because they employ quantities such as bond lengths and bond angles. Much work in geometry optimization was done in *Z-matrix* coordinates, which represent a simple nonredundant internal coordinate system, containing just bond lengths, bond angles, and dihedral angles. However, setting up reasonable *Z*-matrices (i.e., ones with low couplings) can be quite complicated for large molecules, especially for ring systems, and it cannot easily be automated.

If a good approximation to the Hessian is available (e.g., from molecular mechanics) one can create a set of *local normal coordinates* in the frame of the eigenvectors of that Hessian. The advantage of these coordinates is their simple automatic generation and low harmonic couplings. Local normal coordinates are linear combinations of bond lengths and angles and, therefore, do not provide the same vivid picture to the chemist as individual coordinates like *Z*-matrices.

Curvilinear internal coordinates have been used in the earliest gradient geometry optimizations, but the effort needed to set them up limited their usage. However, the superior convergence they can potentially provide has shifted more attention toward them recently.^{13,17,18} The "*natural internal coordinates*,"¹⁶ introduced long ago, minimize both harmonic and anharmonic couplings. They resemble the coordinates used by vibrational spectroscopists³¹ and are built from bond stretchings and linear combinations of bond angles and torsional angles, employing local pseudosymmetry around each atomic center. Their construction is quite complicated but automatic programs have recently become available for their construction.^{17,32} In certain cases it is difficult to avoid using more than $3N - 6(5)$ natural internal coordinates. Such redundancies can be handled by a generalization of the usual optimization procedures.¹⁹ Ultimately, one can use a heavily redundant full set of bond lengths, bond angles, torsional angles, and out-of-plane displacements,^{18,20} avoiding the construction of the natural internals. This method has not been tested in the present work. It is important to ensure that the definition of the internal coordinates stays constant in the course of the optimization. In the present work, we always use the coordinates determined at the starting geometries.

To compare optimizations in different coordinate systems on equal footing, one must use approximations equivalent to the Hessian matrix. Our program transforms a given gradient and Hessian from Cartesian coordinates to either *Z*-matrix coordinates, local normal coordinates, or natural internal coordinates. In most cases, the model Hessian of Lindh et al.¹⁴ was used. The Hessian update was performed in internal coordinates (i.e., after the transformation was carried out).

Results and Discussion

The methods outlined above have been incorporated into the MOLPRO³³ package of *ab initio* programs. To compare the methods systematically, we used a test suite proposed by Baker.¹³ This test suite comprises 30 molecules with 2 to 81 internal degrees of freedom. All optimizations were carried out at the restricted Hartree–Fock (RHF) level employing the STO-3G basis set. Initial Cartesian geometries were taken from Ref. 13. The units for

internal coordinates are atomic units (a_0) for stretchings and radians for angles. Units for energy are atomic units (E_h). Unless noted otherwise, the magnitudes of the components in the optimization step were restricted to 0.3 (a_0 or radians) or less in internal coordinates. The optimization was terminated when the maximum gradient component in internal coordinates was less than 0.0003 (E_h/a_0 or E_h/rad) and either an energy change from the previous cycle was less than $10^{-6} E_h$ or all compo-

nents of the optimization step in internal coordinates were smaller than 0.0003. These convergence criteria were also used by Baker¹³ and Lindh et al.¹⁴

COORDINATE SYSTEMS

Table I shows the number of iterations needed to optimize the equilibrium geometries of Baker's test set of 30 molecules.¹³ For comparison, this

TABLE I. Comparison of Number of Steps Required to Optimize Equilibrium Geometries.

Molecule ^a	Baker ^b	Lindh et al. ^c		LNC ^d		Nat. internal ^d	
	EF	RF	DIIS	RF	DIIS ^e	RF	DIIS ^e
Water	5	4	4	4	4	4	4
Ammonia	6	5	7	5	5	6	6
Ethane	4	4	4	4	4	4	4
Acetylene	6	5	5	5	5	6	6
Allene	5	5	4	5	5	4	5
Hydroxysulfane	11	8	11	7	9	7	7
Benzene	4	3	3	3	3	3	3
Methylamine	5	5	5	5	5	5	5
Ethanol	6	5	5	5	5	5	5
Acetone	7	5	5	5	5	5	5
Disilylether	10	11	11	10	12	9	9
1,3,5-Trisilacyclohexane	8	8	8	8	8	6	6
Benzaldehyde	6	5	5	5	5	5	5
1,3-Difluorobenzene	5	5	5	5	5	5	5
1,3,5-Trifluorobenzene	5	4	4	4	4	4	4
Neopentane	5	5	4	4	4	4	4
Furan	8	7	6	6	7	6	6
Naphtalene	5	6	6	6	6	6	6
1,5-Difluoronaphtalene	6	6	6	6	6	6	6
2-Hydroxybicyclopentane	15	10	8	9	9	9 ^f	9 ^f
ACHTAR10	11	8	8	8	8	9	8
ACANIL01	7	8	8	8	7	8	8
Benzidine	10	10	10	8	8	7	8
Pterin	9	9	9	9	8	9	9
Difuropyrazine	8	7	7	7	7	7	7
Mesityloxide	7	6	5	5	5	6	6
Histidine	30	20	44	23	28	14	14
Dimethylpentane	9	10	9	10	10	10	9
Caffeine	10	7	7	7	7	7	7
Menthone	14	14	14	13	14	10	10
Total	240	215	237	209	218	196	196

^a Starting geometries given in Ref. 13.
^b Ref. 13 using a molecular mechanics Hessian in natural internal coordinates and eigenvector-following (EF) optimization algorithm,⁹ which is very similar to the RF method.
^c Ref. 14 using their model Hessian with BFGS update and local normal coordinates.
^d This work, using the model Hessian of Lindh et al.¹⁴ with BFGS update. LNC = optimization in local normal coordinates; Nat. internal = optimization in natural internal coordinates.
^e Geometry DIIS algorithm (interpolation in the subspace of the gradients).
^f Converged to higher final energy structure than Ref. 13 ($E = -265.46237035 E_h$).

table also includes the best results of Baker and Lindh et al.¹⁴ For each molecule, four calculations are presented: minimization with the rational function or the geometry DIIS methods in either local normal or natural internal coordinates. In all cases, the Hessian matrix was approximated by the model Hessian of Lindh et al.,¹⁴ which was updated in internal coordinates using the standard BFGS procedure.

Using natural internal coordinates, one molecule (2-hydroxybicyclopentane) converged to a higher local energy minimum (*endo* form) than in Ref. 13 (*exo* form). The optimizations in local normal coordinates were essentially the same as in Lindh's calculations, with minor differences, due to different restrictions on the step length. Table I shows that, measured by the total number of iteration steps, natural internal coordinates perform somewhat better than local normal coordinates.

Table II compares optimization in Z-matrix coordinates with local normal and natural internal coordinates for eight medium-sized cyclic molecules, used by Schlegel³⁴ to compare Z-matrix, Cartesian, and mixed coordinate optimizations. The Z-matrix coordinates are stated to have minimal couplings. Nevertheless, Table II shows that Z-matrix coordinates are inferior to the other two coordinate systems. Overall, natural internal coordinates show the best performance. As expected, the performance of all coordinate systems is similar in those cases where the starting geometry is

good, as shown by the low number of iterations (≤ 10) needed for convergence. In the more difficult cases, the differences are larger.

APPROXIMATIONS TO THE HESSIAN MATRIX

Table I shows that the use of Lindh's model Hessian, which depends on the geometry and is recomputed in every iteration, is superior to calculating the Hessian matrix once by a molecular mechanics force field and using it throughout the optimization.¹³ The total number of optimization cycles (196 for RF method) was considerably smaller than in previous work (240 steps using the eigenvector method).¹³

Different updating schemes are compared in Table III. It is readily seen that updating improves convergence rapidly. This is especially true in local normal coordinates, which are more strongly coupled than the natural internal coordinates, and thus depend more critically on the approximation of the Hessian matrix. For all four cases, the standard BFGS update²⁴⁻²⁷ performed much better than the conjugate gradient (CG) update.²³ The problem of the CG update seems to be the reordering of the optimization steps by decreasing energy. If an optimization step leads to an increase in the energy it is not used in the update. So, no new information is incorporated in the Hessian, leading in some cases to infinite oscillation between two geometries. Although, in most cases, the CG up-

TABLE II. Comparison of Number of Steps Required to Optimize Equilibrium Geometries Using Various Coordinate Systems.^a

Molecule	Z-matrix ^b		Local normal		Nat. internal	
	RF	DIIS	RF	DIIS	RF	DIIS
2-Fluorofuran	7	7	7	7	6	6
Norbornane	13	13	6	5	5	5
Bicyclo[2,2,2]octane	16	21	13	15	12	13
Bicyclo[3,2,1]octane	6	6	6	6	6	6
2-Hydroxybicyclopentane	7	8	9	9	8	8
ACTHCP	40	40	24	24	29	22
Histamine H ⁺	40	> 50	34	> 50	21	23
1,4,5-Trihydroxyanthroquinone	8	8	7	7	7	7
Total	142	> 153	106	> 123	94	91

^a Algorithms and parameters as used in Table I. The ">" signs indicate that some molecules failed to converge within 50 optimization cycles (the maximum allowed).

^b Starting geometries given in Ref. 34.

TABLE III.
Total Number of Optimization Cycles in RF Minimizations Using Different Updating Schemes to the Hessian Matrix.^a

Update	Local normal		Nat. internal	
	RF	GDIIS	RF	GDIIS
Diag. ^b	—	—	219	222
None	> 329	250	300	218
Conj. grad. ^c	> 328	> 406	> 314	216
BFGS	209	218	196	196

^a Update was restricted to the use of five geometries and gradients. Unless noted otherwise, the Hessian matrix was approximated by the model Hessian of Lindh et al.¹⁴ The ">" signs indicate that some molecules failed to converge within 50 optimization cycles (the maximum allowed).

^b Hessian matrix approximated by diagonal force field without update. Only available in natural internal coordinates (see Ref. 16).

^c Conjugate gradient update of Schlegel (see Ref. 23).

date behaves similarly to the BFGS, this oscillatory behavior occurring for a few molecules led to its poorer overall performance.

It is wise to restrict the number of geometries included in the BFGS update to prevent the use of information from steps too far away from the actual point which may be misleading. We found the inclusion of the last five geometries into the updating procedure to be a good compromise. Although, in some cases, the inclusion of more geometries leads to faster convergence, there are cases (e.g., molecules with very flexible modes, and therefore large step sizes) in which the inclusion of too many geometries is dangerous.

The algorithm which sets up the natural internal coordinates also provides diagonal force constants as a guess to the Hessian matrix¹⁷ using a simple valence force field.³¹ As seen in Table III, this approximation worked quite well in spite of

its simplicity and that it is not updated—probably because it is tailored to the natural internal coordinates.

LIMITING THE OPTIMIZATION STEP

Table IV shows the performance of RF optimizations in different coordinate systems using various methods of restricting the step length. These procedures were activated if the step norm exceeded the threshold of 0.3 (a_0 or radians). Lower thresholds always slowed down convergence. With higher thresholds the optimization may show oscillative behavior (see below). Unexpectedly, the best performance was achieved with the simplest step limiting method, by cutting back those components of the step vector that exceeded the threshold. Step scaling, which affects all components (scaling of the step norm or the dynamical procedure), showed a slightly worse performance, indicating that if some components of the step vector are too large, the rest of them may still be reasonable and should not be scaled. In some cases, convergence was best without any restriction on the step size. Nevertheless, we do not recommend omitting step restriction, because the Hessian may occasionally have a very small eigenvalue, leading to very large steps and thus far overshooting the minimum.

LINE SEARCH

Combination of the RF method with a line search procedure did not improve convergence of the optimizations—in natural internal coordinates they converged even slower (see Table IV). The fact that additional interpolation procedures do not yield improved performance of geometry optimizations indicates that the RF method generally provides a

TABLE IV.
Total Number of Optimization Cycles in RF Minimizations with Model Hessian¹⁴ and BFGS Update Using Different Step Scaling and Line Search Methods.

Coordinates	Line search	Step scaling			
		None	Stepmax	Stepnorm	Dynamic
Local normal	None	209	209	212	224
Local normal	Partial	209	210	211	214
Natural internal	None	200	196	198	202
Natural internal	Partial	206	206	209	210

reasonable step toward the minimum and does not need further improvement (except for simple step scaling).

GEOMETRY DIIS

Table V compares GDIIS optimizations with different definitions of the error vector. Gradients attach more weight to stiff degrees of freedom, whereas estimated errors in the geometry, $-\mathbf{H}_k \mathbf{g}_k$, weigh floppy coordinates more strongly. Using estimated energy lowerings (i.e., the scalar products of geometry steps and gradients in the DIIS matrix) Eq. (11) is intermediate between these. Using Baker's test suite, which includes both stiff and flexible molecules, the best overall performance was achieved using gradients.

The number of error vectors included in the DIIS procedure is also important. Error vectors which are too far away from the actual geometry are not linear in the geometry and, by contributing misleading information, slow down the optimization. On the other hand, the quality of the DIIS interpolation improves with the number of vectors included. We found the use of the gradients of the last five geometries to be a reasonable compromise which yielded the best convergence for most of the test cases.

We have also compared two methods for the relaxation of the geometry (i.e., the optimization step which is made using the interpolated geometry and gradient). A simple quasi-Newton step, proposed originally,¹⁰ was found to be less effective than a rational function (RF) step (see Table V).

To summarize, the most effective GDIIS interpolation was obtained with a DIIS matrix constructed from, at most, five gradient vectors of the previous geometries. The new geometry was then

predicted using the RF method from the interpolated geometry and gradient vectors. Using natural internal coordinates, the RF and DIIS methods performed about equally well, whereas, in local normal coordinates, the RF method appears to be superior (cf. Tables I and II). On the other hand, Table III shows that the GDIIS method is much less sensitive to the Hessian update procedure. One may therefore argue that GDIIS is the more stable optimization technique.

CONVERGENCE CRITERIA

All calculations reported so far were carried out using Baker's convergence criterion.¹³ Baker uses the maximum component of the gradient together with the maximum component of the step or the energy decrease between the previous two geometries (see left scheme in Fig. 1). This criterion can handle rigid systems, in which small changes in coordinates cause strong changes in the energy (convergence due to small gradient and step) as well as flexible molecules, which show relatively large displacements with small (chemically insignificant) energy changes (convergence due to small gradient and energy change).

We propose another scheme, which is very close to Baker's criterion. It provides virtually the same energy (in all optimizations deviations were below $10^{-7} E_h$), but often saves a gradient calculation as compared with Baker's scheme. We simply check for convergence after the energy calculation, but before gradient evaluation (see right scheme in Fig. 1). As seen in Table VI, this reduces the total number of optimization cycles from 196 to 185 for the rational function optimization in natural internal coordinates using a BFGS-updated model Hessian. The somewhat weaker convergence criterion used in Gaussian^{18,35} reduces the total number of

TABLE V. Total Number of Optimization Cycles in Geometry DIIS Minimizations with Model Hessian and BFGS Update Using Different Approximations to the DIIS Matrix.

Coordinates	Gradients $\mathbf{g}_i^\dagger \mathbf{g}_j$	Scaled energy ^a $\mathbf{g}_i^\dagger \mathbf{H}_k^{-1} \mathbf{g}_j$		Energy ^b $\mathbf{g}_i^\dagger \mathbf{H}_k^{-1} \mathbf{g}_j$	Steps $\mathbf{g}_i^\dagger \mathbf{H}_k^{-1} \mathbf{H}_k^{-1} \mathbf{g}_j$
	RF ^c	QN ^d	RF ^c	RF ^c	RF ^c
Local normal	218	284	218	219	233
Natural internal	196	219	197	197	205

^a Only diagonal elements of the actual Hessian scaled to $0.5 \leq H_{ii} \leq 3.0$ [aJ/Å² or aJ/rad²] were used.

^b All elements of the actual Hessian were used.

^c Relaxation: rational function step.

^d Relaxation: quasi-Newton step.

TABLE VI.
Comparison of Total Number of Steps Required to Optimize Equilibrium Geometries Using Various Convergence Criteria.

	Baker ^a	Lindh ^b	Schlegel ^c	MOLPRO ^d
Baker criterion	240	215	—	196
MOLPRO criterion	—	—	—	185
Gaussian criterion	—	199	183	183

^a Ref. 13: EF optimization in natural internal coordinates using a molecular mechanics Hessian.

^b Ref. 14: RF optimization in local normal coordinates using the model Hessian.

^c Refs. 18 and 35: restricted quasi-Newton step in redundant internal coordinates using an empirical force-field Hessian.

^d This work: RF optimization in natural internal coordinates using the model Hessian of Lindh et al.¹⁴

cycles further to 183, but, in some cases, the error in the energy is much larger (up to $10^{-4} E_h$), so that the final energies given in Ref. 13 are not always reproduced. As shown in Table VI, using the criterion in Gaussian we need the same total number of optimization cycles (183) as Ref. 18. However, using our criterion increases the number of cycles only insignificantly, to 185, and improves the accuracy of the calculated energies.

Conclusions

Calculations for a wide range of molecules demonstrate that the use of curvilinear natural internal coordinates^{16,17} significantly improves the convergence and stability of geometry optimizations. Local normal coordinates defined by the eigenvectors of the Hessian also perform well if a reasonable approximation of the Hessian matrix is available. We used the model Hessian of Lindh et al.,¹⁴ which is recalculated at every optimization step and improved by the BFGS update procedure. These choices of coordinates and Hessian involve no user input and work as "black box" procedures as required for geometry optimizations of large molecules. Of the optimization methods investi-

gated, the rational function (RF) algorithm^{8,9} and geometry DIIS methods¹⁰ performed best. The RF method was found to be more sensitive to the Hessian update procedure, whereas the DIIS method was somewhat more sensitive to the choice of coordinates.

The present results indicate that the optimization algorithms and the coordinate systems do not seem to leave much room for further improvement. Given the simplicity of the model Hessian of Lindh et al.,¹⁴ it performs remarkably well, but further speed-ups may be possible using more sophisticated approximations to the Hessian.

Acknowledgments

We thank Roland Lindh for stimulating discussions and for providing a program that computes the model Hessian. P. P. gratefully acknowledges support by the Air Force Office of Research, the National Science Foundation, and the Alexander-von-Humboldt Foundation.

References

1. P. Pulay, *Adv. Chem. Phys.*, **69**, 241 (1987).
2. Y. Yamaguchi, Y. Osamura, J. D. Goddard, and H. F. Schaefer III, *A New Dimension to Quantum Chemistry—Analytic Derivative Methods in Ab-Initio Molecular Electronic Structure Theory*, Oxford University Press, Oxford, 1994.
3. P. Jørgensen and J. Simons, *Geometrical Derivatives of Energy Surfaces and Molecular Properties*, Reidel, Dordrecht, 1981.
4. H. B. Schlegel, *Adv. Chem. Phys.*, **67**, 249 (1987).
5. J. D. Head, B. Weiner, and M. C. Zerner, *Int. J. Quantum Chem.*, **33**, 177 (1988).
6. J. D. Head and M. C. Zerner, *Adv. Quantum Chem.*, **20**, 239 (1989).
7. R. Fletcher, *Practical Methods of Optimization*, Wiley, Chichester, 1987.

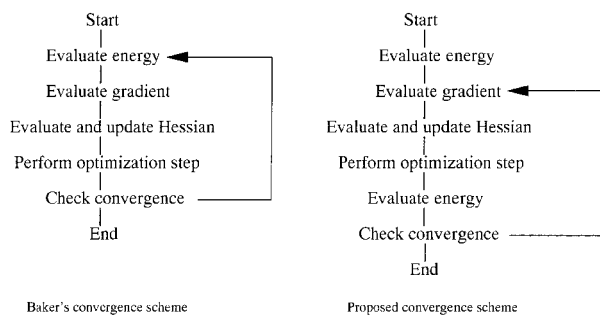


FIGURE 1. Convergence schemes.

8. A. Banerjee, N. Adams, J. Simons, and R. Shepard, *J. Phys. Chem.*, **89**, 52 (1985).
9. J. Baker, *J. Comput. Chem.*, **7**, 385 (1986).
10. P. Császár and P. Pulay, *J. Mol. Struct.*, **114**, 31 (1984).
11. H. B. Schlegel, *Theor. Chim. Acta*, **66**, 333 (1984).
12. T. H. Fischer and J. Almlöf, *J. Phys. Chem.*, **96**, 9768 (1992).
13. J. Baker, *J. Comput. Chem.*, **14**, 1085 (1993).
14. R. Lindh, A. Bernhardsson, G. Karlström, and P. Å. Malmqvist, *Chem. Phys. Lett.*, **241**, 423 (1995).
15. J. Baker and W. J. Hehre, *J. Comput. Chem.*, **12**, 606 (1991).
16. P. Pulay, G. Fogarasi, F. Pang, and J. E. Boggs, *J. Am. Chem. Soc.*, **101**, 2550 (1979).
17. G. Fogarasi, X. Zhou, P. W. Taylor, and P. Pulay, *J. Am. Chem. Soc.*, **114**, 8191 (1992).
18. C. Peng, P. Y. Ayala, H. B. Schlegel, and M. J. Frisch, *J. Comput. Chem.*, **17**, 49 (1996).
19. P. Pulay and G. Fogarasi, *J. Chem. Phys.*, **96**, 2856 (1991).
20. J. Baker, A. Kessi, and B. Delley, *J. Chem. Phys.*, **105**, 192 (1996).
21. P. Jørgensen, P. Swanstrøm, and D. L. Yeager, *J. Chem. Phys.*, **78**, 347 (1983).
22. H.-J. Werner, *Adv. Chem. Phys.*, **69**, 1 (1987), and references therein.
23. H. B. Schlegel, *J. Comput. Chem.*, **3**, 214 (1982).
24. C. G. Broyden, *J. Inst. Math. Appl.*, **6**, 76 (1970).
25. R. Fletcher, *Comput. J.*, **13**, 317 (1970).
26. D. Goldfarb, *Math. Comp.*, **24**, 23 (1970).
27. D. F. Shanno, *Math. Comp.*, **24**, 647 (1970).
28. R. Fletcher and C. M. Reeves, *Comput. J.*, **13**, 317 (1970).
29. D. Lu, M. Zhao, and D. G. Truhlar, *J. Comput. Chem.*, **12**, 376 (1991).
30. J. Baker, *J. Comput. Chem.*, **13**, 241 (1992).
31. E. B. Wilson Jr, J. C. Decius, and P. C. Cross, *Molecular Vibrations*, McGraw-Hill, New York, 1955.
32. R. Ahlrichs, M. Bär, M. Ehrig, M. Häser, H. Horn, and C. Kölmel, *TURBOMOLE, v. 2.1 beta*, Biosym Technologies, San Diego, CA, 1992.
33. MOLPRO is a package of *ab initio* programs written by H.-J. Werner and P. J. Knowles with contributions from J. Almlöf, R. D. Amos, M. J. O. Deegan, F. Eckert, S. T. Elbert, C. Hampel, R. Lindh, W. Meyer, M. E. Mura, K. A. Peterson, R. M. Pitzer, H. Stoll, A. J. Stone, P. R. Taylor, and T. Thorsteinsson, Version 96.4, University of Birmingham, UK, 1996.
34. H. B. Schlegel, *Int. J. Quantum Chem. Symp.*, **26**, 243 (1992).
35. M. J. Frisch, G. W. Trucks, H. B. Schlegel, P. M. W. Gill, B. G. Johnson, M. A. Robb, J. R. Cheeseman, T. Keith, G. A. Petersson, J. A. Montgomery, K. Raghavachari, M. A. Al-Laham, V. G. Zakrzewski, J. V. Ortiz, J. B. Foresman, J. Cioslowski, B. B. Stefanov, A. Nanayakkara, M. Challacombe, C. Y. Peng, P. Y. Ayala, W. Chen, M. W. Wong, J. L. Andres, E. S. Replogle, R. Gomperts, R. L. Martin, D. J. Fox, J. S. Binkley, D. J. Defrees, J. Baker, J. P. Stewart, M. Head-Gordon, C. Gonzalez, and J. A. Pople, *Gaussian-94, Revision D.1*, Gaussian, Inc., Pittsburgh, PA, 1995.